

What Do We Know? A Bibliometric Analysis of Current Literature on COVID-19 and its Implications

Pratyush Yadav
pratyushyadav.yadav@gmail.com

Anju R
anjurvdn@gmail.com

Nargis Pervin
nargisp@iitm.ac.in

Department of Management Studies
Indian Institute of Technology, Madras

Abstract

The COVID-19 pandemic has plunged the world into chaos by affecting people's lifestyles and imposing immense pressures on healthcare professionals. Since its outbreak in Wuhan, China, back in December 2019, researchers all across the globe have been working tirelessly to provide reliable insights to understand and combat the virus. As a result, the number of publications related to the novel coronavirus has been increasing rapidly. This study aims to quantify and summarize the progress of SARS-CoV-2 related research from November 2019 onwards to January 2021 by employing a bibliometric analysis and topic modelling approaches. A total of 33,159 research publications, downloaded from the Web of Science (WoS) core collection database, were analyzed. The key aspects of our study include identifying important publications, their distribution across countries and organizations, important journals and central authors who have made a significant contribution to the current literature. We have also delineated the major themes addressed in the academic community.

1. Introduction

The novel coronavirus disease (COVID-19) has spread at an alarming rate since the first case reported in Wuhan, China, in December 2019, causing people across the globe to confront an unprecedented crisis [1]. The World Health Organization (WHO), on 30th January 2020, announced the outbreak of the virus as a public health emergency of international concern, further declaring it a pandemic on 11th March 2020 [2]. The pandemic has not only caused difficulties in the healthcare sector placing immense pressure on healthcare professionals, but has affected almost all sectors of life, with many countries resorting to lockdowns [3]. The scientific community, in response, has been attempting to gain a better understanding of the virus, its structure, possible

treatments, and potential implications. The sheer volume of research articles in this domain has increased monumentally, spanning across different countries and organizations. Additionally, the faster acceptance and reduced publication times have aggravated several studies on the virus from the research community [4].

While the current era of digital libraries makes the vast and diverse information available online, it inherently poses significant challenges in accessing relevant information and updating our knowledge regarding the virus. As a result, a comprehensive overview of the current status of research is crucial to manage and organize the extensive literature, examine the research areas, identify the key stakeholders and find necessary gaps for the scientific community to work on further. Bibliometric analysis, a statistical approach to analyze citation data, is used for identifying important publications, key authors and journals contributing to coronavirus research. This analysis method has been extensively used in various fields to measure research prowess quantitatively and qualitatively [5, 6].

Bibliometric studies on COVID-19 usually fall under two broad categories: generic (focusing on human coronaviruses) and specific (emphasis on the current crisis). Generic studies talk about human coronaviruses from a historical perspective. These studies intend to provide a holistic viewpoint and aid researchers in learning from previous research and getting directions to face the current crisis. They employ a mixture of descriptive analysis detailing the research prowess and science mapping to identify themes [4, 6]. Although they utilise a comprehensive corpus of articles, they fail to uncover themes or areas pertaining to the current crisis, guiding researchers in the appropriate directions. Specific studies focus on the current crisis the world is experiencing through different perspectives. In some studies there is more emphasis on evaluating research developments and contributions across countries, institutions, authors and journals [7, 8, 9], while others attempt to strike a balance between descriptive evaluations and the identification of intricate themes in the research community [10, 11, 12, 13]. The

work done in these studies provide a glance at the trends and themes in COVID-19 related research; however, they are limited to the initial periods of the pandemic, and the themes discovered are more distributed along the lines of the virus itself. It is also worth mentioning that the citation networks and author collaboration networks are under-utilised to identify the central nodes, which might differ from the rankings based on citations. Moreover, research on COVID-19 is increasing at an alarming rate, making it important to conduct a bibliometric study at different periods to find additional themes discussing the impact and non-medical aspects of the virus, which have gained prominence after the first half of the pandemic's lifetime. The current study aims to decipher the trends and overarching themes inherent in the literature concerning the COVID-19 pandemic.

The scope of this study spans an extensive set of research articles published between November 2019 and January 2021 covering a huge part of the virus' lifetime. Bibliometric analysis aids in the process by identifying important milestones and key contributors in the research community. The analysis also incorporates the influence of nodes while ranking papers and authors in the associated network. Although co-occurrence networks have been used for finding common topics, the network sparsity and size becomes challenging to handle with a large corpus making other approaches more efficient. Therefore, to identify the significant themes, abstracts of research articles were represented using Word2Vec embeddings and similar articles were grouped using clustering-based methods. The following set of research questions are addressed in this study:

RQ1: What is the distribution of articles across countries, organizations, journals and months? Which are the central articles in the citation network?

RQ2: Who are the influential authors in this field and what is their trend of collaboration? Are there more collaborations within the same set of researchers or between diverse groups?

RQ3: What are the significant themes that have emerged from recent research developments?

2. Methodology

2.1. Data Retrieval

The data for this study was retrieved from the Web of Science (WoS) core collection database, one of the major contributing sources for bibliometric data. In order to get relevant articles, search terms including "COVID-19", "2019 Novel Coronavirus",

"SARS-CoV-2", etc. were used. The complete information of articles was exported as a plain-text document on January 22, 2021, comprising fields such as abstracts, keywords, publication venues, author affiliations, cited references, citation counts, etc. The complete query used for obtaining the data is given in Figure 1.

```
TS = (((COVID-19) OR (NOVEL CORONAVIRUS) OR (2019-nCov) OR
(SARS-CoV-2) OR (CORONAVIRUS DISEASE 2019) OR (SARS2) OR
(CORONAVIRUS 2) OR (2019 NOVEL CORONAVIRUS) OR
(CORONA VIRUS DISEASE 2019)))
Refined by: DOCUMENT TYPES: (ARTICLE OR MEETING ABSTRACT OR
PROCEEDINGS PAPER OR LETTER OR NEWS ITEM OR REVIEW
OR EARLY ACCESS OR BOOK CHAPTER)
Timespan: 2019-2021. Indexes: SCI-EXPANDED, SSCI, A&HCI.
```

Figure 1. WoS Data Query

2.2. Data Preparation

Based on the query, a total of 66,607 publications were identified from the WoS core collection. Proper care was taken to exclude duplicate entries, articles with incomplete information (missing abstract or publication date) from the total count. The time period considered for the study further reduced the overall article count to 33,191 entries. Out of these, there were 647 articles that were indexed in WoS under different languages. However, only the articles whose abstract and title was in a language other than English were removed (32 such articles were identified). A final set of 33,159 articles were used for analysis.

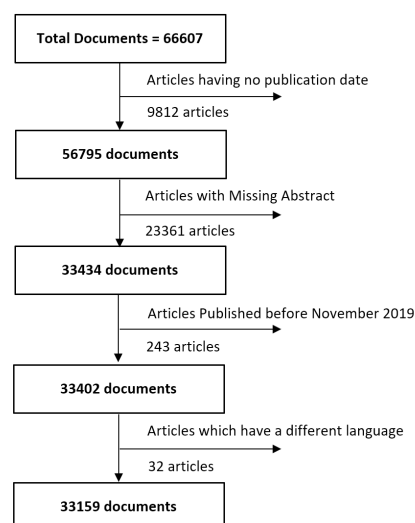


Figure 2. Data Preparation Process

Out of the 33,159 articles chosen, 24,822 (74.85%) were research articles, 5,323 (16.05%) were review

articles, 2,267 (6.83%) constituted editorial materials and the remaining publications fell into the category of letters, meeting abstracts, book reviews, etc. It was also observed that the number of papers falling under the Information Systems (IS) domain were only 220 (0.66%). The majority of the papers in our dataset were published in the year 2020 (94.94%), which seems reasonable since core research on the virus picked up momentum in that year. Although papers from 2021 are present, they constitute a small portion of the dataset spanning 1,618 articles. This is mainly because the data was downloaded in January 2021, allowing us to retrieve a small sample of published articles and a few early access papers. The subsequent subsection details the analysis procedure adopted for our study.

2.3. Method Overview

Primary analysis for the study was carried out using Python v3.7.4 and Tableau was used for visualization purpose. Gephi 0.9.2 was also used for analyzing the large citation and collaboration graphs. The analysis can be broken into the following four subparts:

- **Publication Trend Analysis:** A detailed exploratory analysis on the different characteristics of research articles in our dataset was performed. This included distribution of the articles over a period of time, across countries, organizations. Additionally focus was given to the major journals and authors contributing to the research prowess made in the current field.
- **Citation Network Analysis:** A citation network is a directed network with papers as nodes and the citation relationship between them as edges. For example, a directed edge from paper A to paper B indicates that paper A has cited paper B. The *Cited References* field in the WoS dataset is utilized in order to create a citation edgelist which is then imported in Gephi [14] for further visualization and analysis. This citation network is then used to identify the key papers in the current literature based on their *centrality* [15].
- **Authors and Collaboration Network Analysis:** Unlike citation networks, collaboration networks are undirected in nature and have authors as their nodes, and an edge represents a collaboration between the two authors. While the two networks have a difference in their structure and interpretation, the process of finding key nodes is very similar. The influential authors were identified from the collaboration network ranked according to the betweenness and degree centrality. Additionally, the top authors based on publication as well as citation counts are reported.

- **Theme Identification:** The abstract of a research article can convey a lot of information about the paper and its corresponding aims. Hence, it seems reasonable to build a topic model on the corpus of abstracts to understand the different and diverse themes presented in the literature. For this purpose, Word2Vec [16], a natural language processing technique based on neural networks, was used to map each word in the corpus to a multidimensional vector. This allows us to create a representation of text (popularly known as word embeddings) such that similar words are closer to each other in the vector space. A Word2Vec model was trained on the 33,159 articles in our dataset, and average word embeddings were used to represent each article [17]. Before training the Word2Vec model, a series of preprocessing steps were carried out on the textual data. The commonly occurring words known as stopwords were removed. Stanford Core NLP API was used to perform part-of-speech (POS) tagging and lemmatization to remove the inflectional endings associated with words and retain only the root word. In addition to unigrams, n-grams were further created and considered during the corpus creation. However, unigrams that are subsets of a particular n-gram phrase were removed from the final corpus. For example, for the bi-gram “mental health”, if the unigrams “mental” and “health” are present in the corpus, they were removed. After training the Word2Vec model, K-means clustering was applied to the vectors, and each article was assigned to a unique cluster. This allowed for cluster profiling in terms of article abstracts and identifying the important themes in this domain.

3. Results and Discussion

We have adopted both descriptive and exploratory analysis to address the aforementioned research questions.

3.1. COVID-19 Publication Analysis

Figure 3 shows the trend of articles published on COVID-19 in the chosen period of study. It is clear that the topic of COVID-19 is gaining constant momentum, and plentiful contributions from the various research areas imply that a large number of researchers are working towards understanding COVID-19. In the year 2019, the number of articles published on the topic of interest were very low. However, there is a drastic increase in the articles published on COVID-19 as we move to 2020, with the highest number of publications recorded in December 2020. While we observed a

declining trend in the number of papers post 2020, it can not be treated as a hint towards the low research interest of COVID-19. This is because these publications were available as early access papers (for example, a paper whose publication date falls in December 2021 but was accessible on a prior date) during the preparation of this manuscript and do not represent the true number of papers published in those specific months.

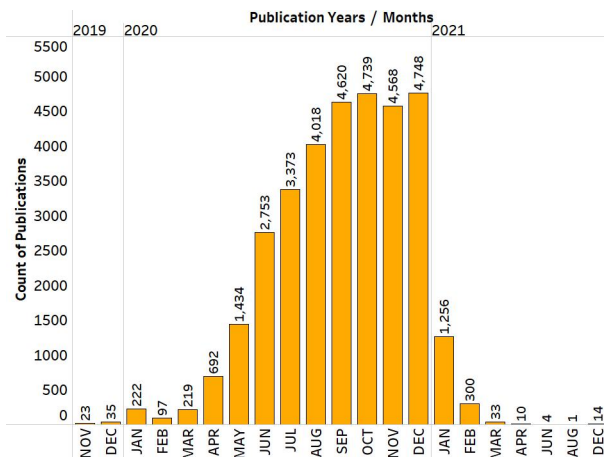
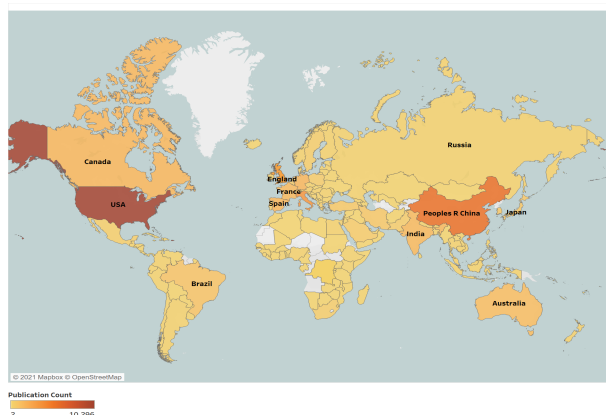


Figure 3. Count of Publications on COVID-19

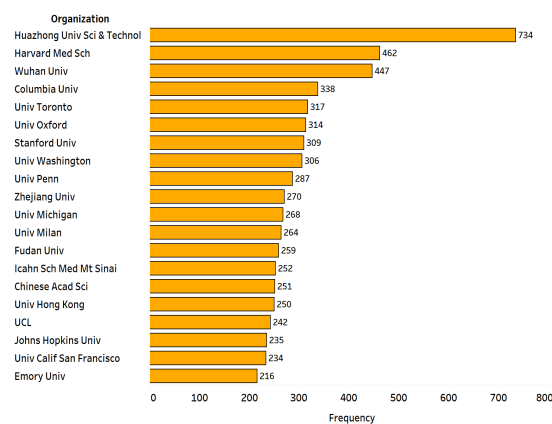
The geographical distribution of the research output in terms of the publication count has been displayed in Figure 4a. This global view indicates the contributions and performance of researchers from each country. The number of publications is represented using the colour intensity in the figure. A darker shade is indicative of a higher contribution from the corresponding countries. According to the number of publications, the top countries are labeled in the figure. While it comes as no surprise that the United States is the leading country in



(a)

terms of publication count, followed by China coming in as the second most contributing country, it is also worth noting the presence of developing countries in the race, such as India, Russia, etc. Similar to the distribution of countries, the top 20 organizations working extensively on COVID-19 research are shown in Figure 4b. “Huazhong University of Science and Technology” tops the list with 734 publications. Harvard Medical School and Wuhan University are in the second and third positions with 462 and 447 publications, respectively. Among the top 20, participation of organizations from China, the United States, Canada, the United Kingdom, Italy, and Hong Kong are notable. Interestingly, two of the top three institutions are located in Wuhan, China, where the first case of the pandemic was reported.

While discussing the publication trends, it is also essential to look at the publication venues where the majority of papers are being published. From the dataset, 3,740 unique journals from various disciplines were found, which is an indication of the diverse research on COVID-19. Based on the total number of publications, the top 10 journals have been listed in Table 1. The International Journal of Environmental Research and Public Health tops the list with 739 publications with an impact factor of 2.849. There are a total of 3,967 citations received towards it during the study period. Among the rest of the journals, in terms of citations, the Journal of Medical Virology is worth mentioning. Journal of Medical Virology is about fundamental and applied research concerning viruses affecting humans. In this journal, there are 333 publications with 10,972 total citations received. While LANCET, The Journal of the American Medical Association, The New England Journal of Medicine are important venues in this field with high citation



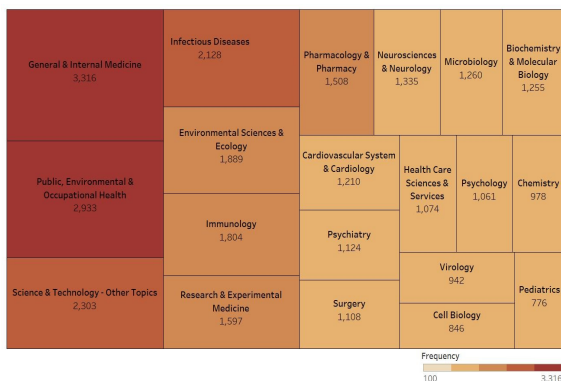
(b)

Figure 4. Publication trends across (a) Countries and (b) Organizations

Table 1. Top 10 Journals with Highest Publication Count

Journal	Total Publications	Total Citations	Impact Factor	Country
International Journal of Environmental Research and Public Health	739	3,967	2.849	Switzerland
PLOS One	666	1,559	2.740	United States
Sustainability	351	478	2.576	Switzerland
Journal of Medical Virology	333	10,972	2.021	United States
Journal of Medical Internet Research	288	751	5.030	Canada
International Journal of Infectious Diseases	270	4,927	3.202	Netherlands
Journal of Clinical Medicine	268	2,741	3.303	Switzerland
Frontiers in Psychology	263	298	2.067	Switzerland
Frontiers in Public Health	256	619	2.483	Switzerland
Science of the Total Environment	249	4,432	6.551	Netherlands

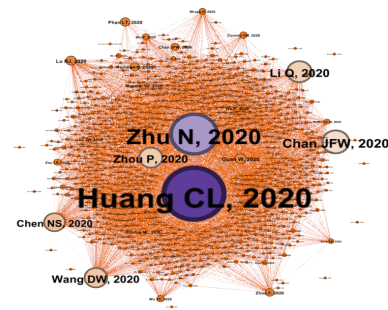
counts, they do not rank among the top in terms of publication counts. The number of publications in the chosen period belong to a total of 151 diverse research areas concerning COVID-19 and the distribution of the top 20 research areas is depicted in Figure 5. The most prominent research area consisting of the

**Figure 5. Major Research Areas**

highest number of publications belong to General and Internal Medicine with a total of 3,316 articles. As can be seen in the figure, most research areas deal with a specified field of medicine or biological sciences (like Immunology, Infectious Diseases, Cardiovascular System & Cardiology). However, it is also interesting to find diverse areas in the form of Psychology (1,061 papers), Science & Technology (2,303 papers) and Environmental Sciences and Ecology (1,889 papers). This gives us a broad idea of the diversified studies that have been conducted on COVID-19.

So far, we have analyzed the publication trends and distributions across countries and publication venues. However, it is also imperative to find the most important and prestigious papers out of all the publications. These

papers become central to understanding the key aspects of the subject area. In line with this thought, the citation network for all publications in the dataset was created as mentioned in Section 2.3. Using this network, the

**Figure 6. Top Papers in Citation Network**

top 5 publications with higher influence were identified based on PageRank. PageRank, an algorithm developed by Google to rank search results, can also be applied in the case of citation networks in identifying the most influential papers. A paper is assigned a PageRank score based on the count and the quality of links pointing to it. Thus, a paper is said to have a higher PageRank score if it is cited by many papers that have a high citation count [18]. These publications can be identified from Figure 6, where the bigger node size represent papers with a higher score. Clearly [19] has the highest influence in the network followed by [20] and [21]. The characteristics of these papers are presented in further detail in Table 2.

3.2. Author Analysis

While discussing the literature on COVID-19, it is also imperative to discuss the authors who have contributed towards the same. Figure 7 shows the top 20

Table 2. Top 5 Publications with Highest Influence in the Network

Article	First Author	Journal	Total Citations	Organization	Country
Clinical Features of Patients infected with 2019 Novel Coronavirus in Wuhan, China	Huang, CL	LANCET	10,453	Jin Yin Tan Hospital, Wuhan	China
A Novel Coronavirus from patients with Pneumonia in China, 2019	Zhu, N	New England Journal of Medicine	5,557	Chinese Centre for Disease Control & Prevention	China
A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster	Chan, JFW	LANCET	2,258	University of Hong Kong	China
Early transmission dynamics in wuhan, china, of novel coronavirus-infected pneumonia	Li, Q	New England Journal of Medicine	3,288	Chinese Centre for Disease Control & Prevention	China
A pneumonia outbreak associated with a new coronavirus of probable bat origin	Zhou, P	Nature	3,986	Chinese Academy of Sciences	China

authors based on publication count and citation count. From the two plots, Figure 7a and Figure 7b, it is clear that the authors in both the rankings are different. While Giuseppe Lippi is the top author based on publication count with a total of 77 publications, the author does not appear in the ranking based on citations. Similarly, Ting Yu has the highest citations (24,480); however, the author does not appear in the other ranking. It is interesting to see that Kwok Yung Yuen is the only author appearing in both the plots; however the ranking is different in both cases. These results bolster the fact that influential authors may not always be the key contributors to scientific literature.

A collaboration network of the authors is shown in Figure 7c. In order to generate this network, the author names were first disambiguated, and each author was assigned a unique identifier. Finally, the edge-list was imported in Gephi, and the network was displayed using the Radial Axis Layout. There are a total of 18,54,182 unique collaborations between authors and 91.4% of these collaborations (16,94,996) occur only once. A filter was applied to the network where two authors having more than five collaborations between them were retained to prioritize authors that collaborate more frequently.

As can be seen from the figure, all author

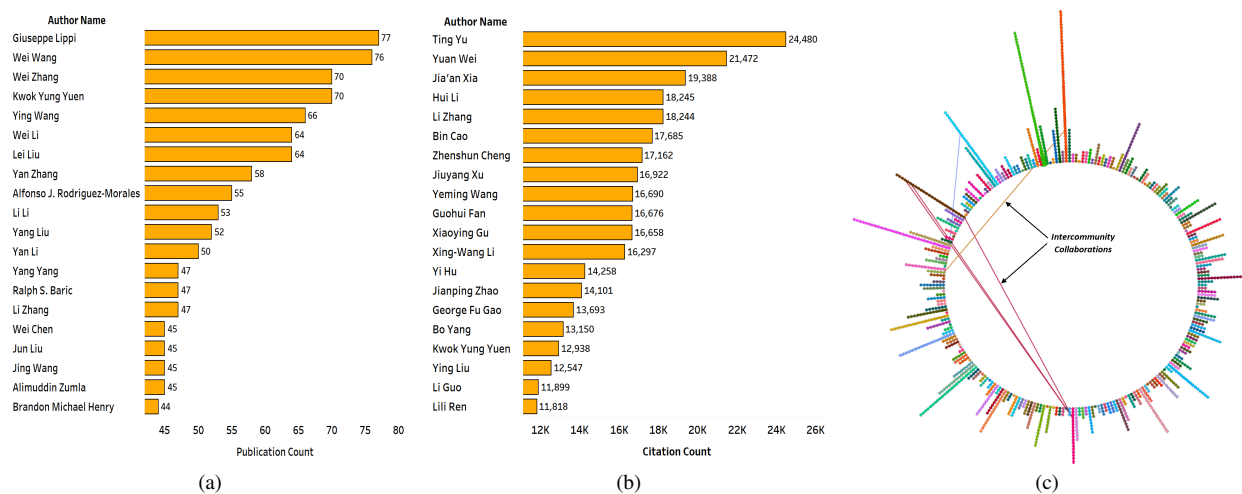


Figure 7. Author Analysis: (a) Top Authors based on Publication Count (b) Top Authors based on Citation Count (c) Author Collaboration Trend

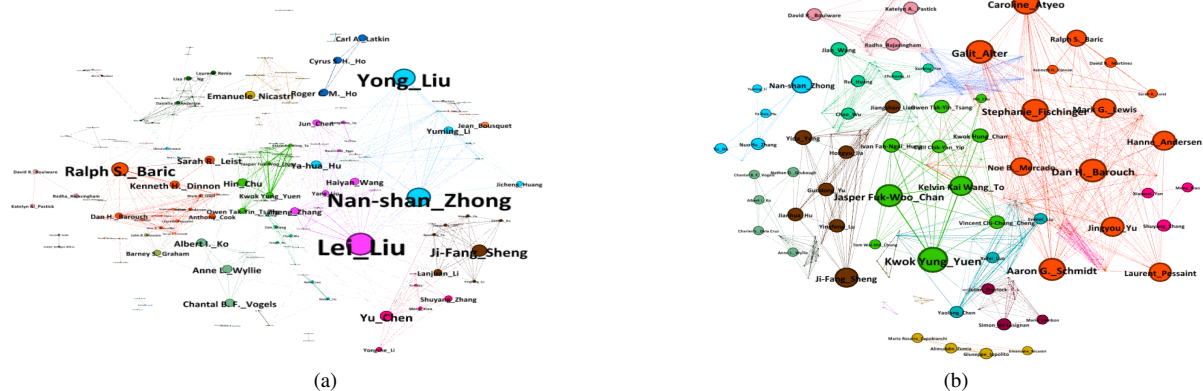


Figure 8. Authors ranked based on (a) Betweenness Centrality and (b) Degree Centrality

communities (obtained from Gephi) are arranged in a circular fashion, while the lines (or edges) indicate the interaction of authors from different communities. Clearly, there are many sub-communities of authors which are strongly connected, indicating a high level of collaboration between them; however, the collaboration between communities is very low. This implies that although several collaborations exist in the research community, strong collaborations exist between the same set of authors, but the diversity of collaborations is limited. This makes the collaboration network a disconnected graph with many connected sub-components. From the collaboration network, influential authors based on centrality measures were also identified. Figure 8a shows the authors who act as bridges towards new collaborations or interactions between different author communities are highlighted. Another interesting point to note is that only a few authors have a high betweenness centrality, which is expected from the collaboration trend and pattern discussed above. Based on the degree centrality, there are multiple authors, for examples Kwok Yung Yuen, Caroline Atyeo, Aaron G. Schmidt, with high rank (Figure 8b).

3.3. Keyword Analysis

Author keywords consist of the important word(s), usually nouns, that in essence, like an abstract, convey information about the different ideas being discussed in the academic literature. In our study, out of 39,123 unique keywords, we have identified the top keywords that authors frequently specify in their papers. We observed that 73.46% of the keywords occur only once. Figure 9 depicts the top 20 keywords based on frequency. However, we have excluded the terms

which are a part of our initial query since it is trivial for them to appear at the top of our list. From Figure 9, it can be seen that, “Mental Health” has the highest frequency followed by other keywords such as “Public Health”, “Pneumonia”, “Anxiety”, “Telemedicine”, “Depression”, “Stress”, “Infectious Disease”, etc. Another key point to note, considering

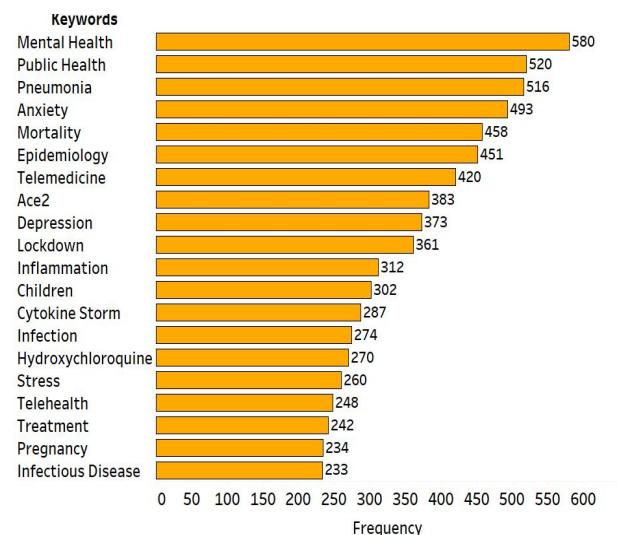


Figure 9. Frequent Author Keywords

the top journals with high publication count, it was observed that the frequently occurring keywords were correlated with the field of the journal. The majority of the frequently occurring terms are related to the mental well-being of the general public, including the healthcare professionals who are facing tremendous pressure owing to the pandemic. On the other hand, while we do have keywords like “Epidemiology” and “Mortality”, they are ranked lower in the list. We believe that this pattern brings light to how the primary topic has

shifted focus from the virus itself to the impact caused by it.

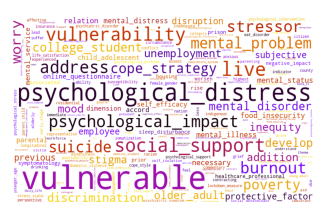
3.4. Identification of Major Research Themes

Based on the preprocessed data and trained Word2Vec model, K-means clustering was carried out on the set of papers (as discussed in Section 2.3). A total of 23 diverse themes were identified; they are listed in Table 3 in decreasing order of the number of papers. The themes were given names according to the papers belonging to them and the corresponding word clouds. From the table, we see that the top 4 themes consist of a total of 17,573 (53%) papers. These themes discuss predominantly the patient care and mortal danger caused during and as a consequence of the COVID-19 situation. They also discuss the process of drug discovery and presence of antiviral defense mechanisms (antibodies) to fight COVID-19. There are certain themes that discuss the virus itself based on its structure, evolution and mutations. These themes contribute to a total of 1,911 publications. Some of the interesting themes are discussed below. These topics can be seen in the word cloud in Figure 10.

1. **Public Well Being during COVID-19:** The primary topics of discussion in this theme include the mental health of the general public, healthcare professionals, burnout and job satisfaction, parental stress and associated coping mechanisms.
2. **Politics, Governance and Health Management:** The pandemic has definitely affected the lives of people, but it has also made it difficult for administrative authorities to manage the issues created and help

mitigate the crisis. This theme deals with papers talking about the government and actions taken by them.

3. **Public Attitudes towards Health Management:** With many countries imposing lockdowns and other strategies to counter the virus, the success of such measures depends vastly on public compliance and willingness to adhere to them. This theme discusses the attitudes of the public towards the virus and preventive behaviours.
4. **Technology Enabled Solutions to Combat COVID-19:** This theme talks about the various technological advancements, especially in medicine, in terms of the novel approaches to detect and curb the virus. It also talks about the use of machine learning and neural networks applied to chest images and others.
5. **Economical Disruptions:** This community talks in detail about disruptions that the virus has brought to various economies. It talks about the effect on the financial sector, tourism and hospitality industries, agricultural sector, airline industry, etc. The dynamics of consumer and firm relationships are also discussed.
6. **The Pandemic and Challenges to Education:** The transition to virtual platforms for learning has presented various challenges to teachers and students alike. This theme expresses those challenges and discusses how smooth the transition has been along with how people are adjusting to the “new normal”.
7. **Effects of Lockdown on Environment:** The pandemic and lockdown have had a profound



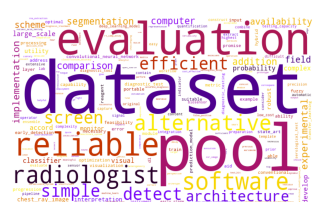
Public Well Being During Covid-19



Politics, Governance and Health Management



Public Attitudes Towards Health Interventions



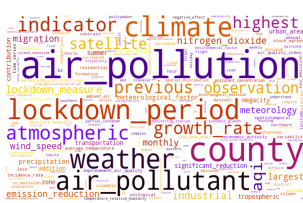
Technology Enabled Solutions to Combat COVID-19



Economical Disruptions



The Pandemic and Challenges to Education



Effects of Lockdown on the Environment



Lifestyle Choices and Habits during Lockdown

Figure 10. Interesting Research Themes

Table 3. Research Themes identified using Word2Vec

Name	Number of Documents	Number of Citations
Patient Care amidst COVID-19	5,297	37,033
Antiviral Defense Mechanisms and Drug Discovery	4,844	77,893
Clinical Conditions, Comorbidities and Fatality	4,719	90,063
Aspects of COVID-19	2,713	20,164
Public Well Being during COVID-19	1,842	13,251
Dynamics of COVID-19 Contagion	1,357	20,226
Structural Characterization and Gene Expression	1,301	13,343
Clinical Manifestations and Complications	1,082	20,095
Politics, Governance and Crisis Management	1,018	5,069
Public Attitudes Towards Health Interventions	956	4,791
Infection Transmission and Control Measures	921	7,183
Technology Enabled Solutions to Combat COVID-19	919	6,675
Economical Disruptions	813	2,657
Clinical Manifestations and Diagnostics	796	34,615
Detection of COVID-19 and Associated Antibodies	715	17,717
Mathematical Modelling	689	8,636
The Pandemic and Challenges to Education	676	1,465
Virus Evolution and Mutation	610	22,113
Effects of Lockdown on Environment	554	5,774
Maternal and Neonatal Care	541	8,918
Patient Management and Safety of Medications	439	4,909
Lifestyle Choices and Habits during Lockdown	240	1,091
COVID-19 and Digital Media	117	421

effect on the environment, especially the air quality. This community discusses the impact on various environmental factors such as reductions in traffic-related black carbon, ultrafine particle number concentrations, etc. The extent of the lockdown being a prime cause for these effects is also discussed.

8. ***Lifestyle Choices and Habits During Lockdown:*** People across the globe have been adapting to the “new normal”. One of the major issues of concern is the impact of the pandemic on health-related quality of life, drinking habits and eating behaviours. These manifest in the form of sedentary behaviour, mental distress, stigmatization due to weight gain, etc. This theme talks about the importance of physical exercise and maintaining a healthy lifestyle.

4. Conclusion

This study aims to uncover the publication trends and themes predominantly discussed in the arena of COVID-19 research. The publications on COVID-19 had witnessed a continuous rise from 2019 to December 2020. The United States has been a significant contributor to the research articles being disseminated

in the academic community; however, other developing countries like China are not that far behind and are making several strides in various research areas including COVID-19. Interestingly, all the authors of the top 5 influential papers belong to China. Huazhong University of Science and Technology, China, has topped the list of organizations in publication count. The International Journal of Environmental Research and Public Health leads in publication counts, whereas the Journal of Medical Virology received the most citations. While General and Internal Medicine was the most prominent research area, it was interesting to find diverse areas like Psychology, Science & Technology, etc. Giuseppe Lippi was the author with the highest number of publications, while Ting Yu received the highest citations. Although there are many collaborations among authors, community analysis on the collaboration network revealed that the interactions between them were minimal. In addition, 23 unique and diverse communities were identified, focusing on the various aspects of the virus and its associated consequences. Our analysis also identified key research themes focused on the pandemic’s impact on education, environment, lifestyle, mental health and administrative policies that provided new insights into the existing

literature.

The current study uses an exhaustive set of articles to identify important trends and generate interesting themes. However, this overall analysis takes a more generic standpoint and aims to be a precursor to further studies specifically identifying the timeline of the shift in research from the virus itself to other associated aspects. We believe that further breaking the themes into sub-themes can highlight the various perspectives at a more granular level and present compelling research opportunities especially the ones discussing the non-medical aspects of COVID-19. For example, the current analysis revealed that the Information Systems (IS) community represents a minor portion of the overall corpus. Thus, the role of the IS community and their contributions so far can be a point of future exploration.

There are few limitations to this study. We have chosen WoS core collection as the principal source of bibliometric data and do not incorporate data from other sources such as Scopus, PubMed, Dimensions, etc. Although we have chosen a vast corpus for analysis, we only consider “English” as the primary language. There might be more articles from other languages, which might further enrich the conclusions.

References

- [1] C. Wang, P. W. Horby, F. G. Hayden, and G. F. Gao, “A novel coronavirus outbreak of global health concern,” *The lancet*, vol. 395, no. 10223, pp. 470–473, 2020.
- [2] J. Hua and R. Shaw, “Corona virus (covid-19) “infodemic” and emerging issues through a data lens: The case of china,” *International journal of environmental research and public health*, vol. 17, no. 7, p. 2309, 2020.
- [3] K. Dube, G. Nhamo, and D. Chikodzi, “Covid-19 cripples global restaurant and hospitality industry,” *Current Issues in Tourism*, vol. 24, no. 11, pp. 1487–1490, 2021.
- [4] Z. Deng, J. Chen, and T. Wang, “Bibliometric and visualization analysis of human coronaviruses: prospects and implications for covid-19 research,” *Frontiers in cellular and infection microbiology*, vol. 10, p. 529, 2020.
- [5] L. Bornmann and R. Mutz, “Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references,” *Journal of the Association for Information Science and Technology*, vol. 66, no. 11, pp. 2215–2222, 2015.
- [6] E. H. Viedma, J. R. L. Robles, J. Guallar, and M. J. Cobo, “Global trends in coronavirus research at the time of covid-19: A general bibliometric approach and content analysis using scimat,” *El profesional de la información*, vol. 29, no. 3, p. 11, 2020.
- [7] R. K. Farooq, S. U. Rehman, M. Ashiq, N. Siddique, and S. Ahmad, “Bibliometric analysis of coronavirus disease (covid-19) literature published in web of science 2019–2020,” *Journal of Family & Community Medicine*, vol. 28, no. 1, p. 1, 2021.
- [8] D. W. Hook, S. J. Porter, H. Draux, and C. T. Herzog, “Real-time bibliometrics: dimensions as a resource for analyzing aspects of covid-19,” *Frontiers in Research Metrics and Analytics*, vol. 5, p. 25, 2021.
- [9] M. Su, S. Xu, and J. Weng, “A bibliometric study of covid-19 research in web of science,” *Pharmacological Research*, 2021.
- [10] Y. Yu, Y. Li, Z. Zhang, Z. Gu, H. Zhong, Q. Zha, L. Yang, C. Zhu, and E. Chen, “A bibliometric analysis using vosviewer of publications on covid-19,” *Annals of translational medicine*, vol. 8, no. 13, 2020.
- [11] J. Wang and N. Hong, “The covid-19 research landscape: measuring topics and collaborations using scientific literature,” *Medicine*, vol. 99, no. 43, 2020.
- [12] A. Aristovnik, D. Ravšelj, and L. Umek, “A bibliometric analysis of covid-19 across science and social science research landscape,” *Sustainability*, vol. 12, no. 21, p. 9132, 2020.
- [13] A. Abd-Alrazaq, J. Schneider, B. Mifsud, T. Alam, M. Househ, M. Hamdi, and Z. Shah, “A comprehensive overview of the covid-19 literature: Machine learning-based bibliometric analysis,” *Journal of Medical Internet Research*, vol. 23, no. 3, 2021.
- [14] M. Bastian, S. Heymann, and M. Jacomy, “Gephi: an open source software for exploring and manipulating networks,” in *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 3, 2009.
- [15] K. Das, S. Samanta, and M. Pal, “Study on centrality measures in social networks: a survey,” *Social network analysis and mining*, vol. 8, no. 1, pp. 1–11, 2018.
- [16] T. Mikolov, K. Chen, G. Corrado, and J. Dean, “Efficient estimation of word representations in vector space,” *arXiv preprint arXiv:1301.3781*, 2013.
- [17] D. Spathis, N. Passalis, and A. Tefas, “Interactive dimensionality reduction using similarity projections,” *Knowledge-Based Systems*, vol. 165, pp. 77–91, 2019.
- [18] N. Ma, J. Guan, and Y. Zhao, “Bringing pagerank to the citation analysis,” *Information Processing & Management*, vol. 44, no. 2, pp. 800–810, 2008.
- [19] C. Huang, Y. Wang, X. Li, L. Ren, J. Zhao, Y. Hu, L. Zhang, G. Fan, J. Xu, X. Gu, *et al.*, “Clinical features of patients infected with 2019 novel coronavirus in wuhan, china,” *The lancet*, vol. 395, no. 10223, pp. 497–506, 2020.
- [20] N. Zhu, D. Zhang, W. Wang, X. Li, B. Yang, J. Song, X. Zhao, B. Huang, W. Shi, R. Lu, *et al.*, “A novel coronavirus from patients with pneumonia in china, 2019,” *New England journal of medicine*, 2020.
- [21] J. F.-W. Chan, S. Yuan, K.-H. Kok, K. K.-W. To, H. Chu, J. Yang, F. Xing, J. Liu, C. C.-Y. Yip, R. W.-S. Poon, *et al.*, “A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster,” *The lancet*, vol. 395, no. 10223, pp. 514–523, 2020.
- [22] A. Abbasi, L. Hossain, and L. Leydesdorff, “Betweenness centrality as a driver of preferential attachment in the evolution of research collaboration networks,” *Journal of Informetrics*, vol. 6, no. 3, pp. 403–412, 2012.